



Autour du lexique et de la phraséologie des écrits scientifiques.

Agnès Tutin

► To cite this version:

Agnès Tutin. Autour du lexique et de la phraséologie des écrits scientifiques. . Revue Française de Linguistique Appliquée, 2007, Autour du lexique scientifique, 12 (2), pp.5-14. hal-01336284

HAL Id: hal-01336284

<https://hal.science/hal-01336284>

Submitted on 22 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AUTOUR DU LEXIQUE ET DE LA PHRASÉOLOGIE DES ÉCRITS SCIENTIFIQUES

Agnès Tutin

Pub. linguistiques | « *Revue française de linguistique appliquée* »

2007/2 Vol. XII | pages 5 à 14

ISSN 1386-1204

Article disponible en ligne à l'adresse :

<http://www.cairn.info/revue-francaise-de-linguistique-appliquee-2007-2-page-5.htm>

Pour citer cet article :

Agnès Tutin, « Autour du lexique et de la phraséologie des écrits scientifiques », *Revue française de linguistique appliquée* 2007/2 (Vol. XII), p. 5-14.

Distribution électronique Cairn.info pour Pub. linguistiques.

© Pub. linguistiques. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

Autour du lexique et de la phraséologie des écrits scientifiques

Agnès Tutin
Université Stendhal-Grenoble 3

Dans les études linguistiques sur les écrits scientifiques, le lexique non terminologique est paradoxalement peu étudié. Les travaux récents portent en effet bien plus fréquemment sur les aspects énonciatifs ou rhétoriques (Fløttum & al. 2006 ; Hyland 1998 ; Poudat 2005 ; Rinck 2006 ; Rinck & al. 2007 ; Grossmann & Wirth 2007) ou sur les différences disciplinaires (Fløttum 2007 ; Hyland & Bondi 2006) que sur les aspects lexicaux à proprement parler. Le lexique a surtout été abordé dans les travaux sur le FOS (Français sur Objectifs Spécifiques) ou dans les travaux anglo-saxons à travers l'*English for Academic Purposes* (par exemple, Coxhead 2000 ou Hyland 2005) dans la perspective de l'enseignement des langues. Pour le français, les recherches approfondies restent assez anciennes – hormis la thèse récente de Pecman (2004) – et les travaux pionniers de Phal (1971), qui demeurent la référence, méritent d'être actualisés. Même si l'entrée lexicale paraît parfois difficile à circonscrire – comment définir et délimiter exactement ce lexique propre aux écrits scientifiques ? –, elle nous semble tout à fait centrale pour une réflexion à la fois épistémologique et linguistique sur les différents types d'écrits scientifiques, que l'on observe la variable disciplinaire ou que l'on cherche à mettre en évidence les divergences entre les genres d'écrits scientifiques (articles, thèses, rapports, manuels). Cette thématique apparaît en outre essentielle pour la linguistique appliquée, tant pour la didactique des langues (aide à la rédaction en langue maternelle, didactique du Français Langue Etrangère pour l'insertion des étudiants étrangers) que pour le Traitement Automatique du Langage dans l'accès à l'information scientifique où des besoins importants apparaissent.

Un lexique de genre, entre terminologie et langue générale

On peut supposer l'existence d'un lexique partagé dans les écrits scientifiques au moins pour deux raisons. D'une part, parce que les écrits scientifiques renvoient à des pratiques intellectuelles qui, malgré les différences inhérentes aux disciplines et aux familles de disciplines, présentent des objectifs et des procédures communs – opérations d'analyse, de raisonnement, d'évaluation – qui seront nécessairement réalisés par des éléments lexicaux dans les écrits. D'autre part, parce que les écrits scientifiques renvoient à des genres codifiés par des communautés de discours, pour reprendre le terme de Swales (1990), qui, par-delà les disciplines, poursuivent peu ou prou les mêmes buts, et donc recourraient à des rhétoriques comparables. La communauté universitaire emploierait donc une langue spécifique – Biber étudie même ces pratiques langagières sous le terme d'*University Language* (Biber 2006)¹ – dans les écrits scientifiques. Le lexique qui nous intéresse serait donc un lexique de genre,

¹ Voir aussi l'étude de Kocourek (1991) sur la langue scientifique et technique.

traversant en grande partie les disciplines. On peut faire l'hypothèse que jusqu'à un certain point, les caractéristiques du lexique non terminologique sont davantage corrélées au sous-genre (article scientifique, vulgarisation ou manuels par exemple) qu'à la discipline même. Les études sur les marqueurs de métadiscours de Hyland (2005) semblent d'ailleurs aller en ce sens. On ne peut cependant nier, comme on le verra dans les contributions d'Averil Coxhead et David Hirsh ainsi que celle de Peter Blumenthal dans ce numéro, l'existence d'un lexique spécifiquement lié à des familles de disciplines (sciences dures, sciences appliquées, sciences humaines) qu'on ne peut pas véritablement considérer comme transdisciplinaire.

Le lexique transdisciplinaire ne renvoie pas aux objets scientifiques des domaines de spécialité, mais au discours sur les objets et les procédures scientifiques. Il peut être considéré de ce point de vue comme un lexique métascientifique (cf. la notion d'opérateur métascientifique dans les travaux de Harris et ses collègues (1989) sur les écrits d'immunologie). Mais le lexique transdisciplinaire des écrits scientifiques ne se cantonne pas au lexique métascientifique. Il inclut aussi un ensemble d'éléments du « métadiscours » au sens de Hyland (2005) ou Dahl (2003), ces marques linguistiques qui ne renvoient pas au sens propositionnel, mais aux interactions au sens large entre un auteur et son destinataire dans une même communauté, et renvoient à des relations internes au discours². Ces éléments de métadiscours incluent les connecteurs, mais aussi le métatexte, les évaluatifs, les marques d'atténuation (*hedges*) ou les marques de prise de position. Les marques du métadiscours scientifique et le lexique métascientifique ne forment pas deux ensembles disjoints. Par exemple, des expressions comme : *méthode prometteuse*, *résultats décevants*, *approche inadaptée* renvoient à la fois à un discours sur l'objet scientifique (de type métascientifique) et un positionnement de l'auteur (métadiscours). En revanche, des expressions comme *élaborer un protocole* ou *collecter des données* renverraient exclusivement au lexique métascientifique, alors que certaines marques métatextuelles comme *comme nous l'examinerons plus loin* relèveraient plus spécifiquement du métadiscours scientifique.

Un lexique difficile à circonscrire

Circonscrire le lexique scientifique général ne va pas de soi. D'une part, parce que l'extension donnée aux corpus étudiés par les chercheurs est variable. D'autre part, parce qu'il est difficile de répartir les mots, par essence polysémiques et polyfonctionnels, dans des strates étanches.

Dans les études sur le lexique scientifique, le terme de « sciences » reçoit des acceptions variables : s'agit-il exclusivement de « sciences dures » comme dans l'étude de Phal sur le *Vocabulaire Général d'Orientation Scientifique* (Phal 1971) ou inclut-on également les sciences appliquées (cf. Coxhead & Hirsh, ce numéro ; Pecman 2004), voire de façon plus large les sciences humaines et sociales (Drouin, ce numéro) ? En outre, les types d'écrits scientifiques examinés pour déterminer ce lexique sont également très divers : certains auteurs se centrent exclusivement sur les écrits de recherche du type thèse ou article scientifique (par exemple, Drouin, ce numéro), d'autres s'intéressent uniquement aux écrits didactiques (Phal 1971) ou à la vulgarisation scientifique (cf. dans ce numéro les contributions de Blumenthal et Poudat & Loiseau).

À côté de la variabilité des supports et des disciplines analysés, reste la difficulté à délimiter ce noyau lexical qu'on considère propre aux écrits scientifiques. De façon

² Voici la définition donnée par Hyland (2005,38) : "[The key principles] are : 1. that metadiscourse is distinct from propositional aspects of discourse; 2. that metadiscourse refers to aspects of the text that embody writer-reader interaction; 3. that metadiscourse refers only to relations which are internal to the discourse".

schématique, on peut imaginer mettre en évidence plusieurs strates lexicales plus ou moins étanches, allant du lexique général à la terminologie du domaine. Pour illustrer concrètement notre propos, nous nous appuyons ici sur un exemple de résumé d'article d'économie.

L'évaluation du coût de l'enfant en coupe instantanée pose le problème de l'hétérogénéité des préférences entre les ménages selon qu'ils ont ou non des enfants. La difficulté théorique de réaliser des comparaisons de bien-être interpersonnelles ne joue cependant plus lorsque l'on dispose de données longitudinales. Celles-ci permettent une observation directe de l'influence de l'arrivée ou du départ des enfants sur le niveau de bien-être tel qu'il est perçu par les ménages. Le panel européen fournit ainsi des éléments subjectifs sur l'aisance financière des agents, mais sous forme de réponse polytomique. Il se pose cependant un problème lié à l'estimation de l'hétérogénéité des comportements lorsque la dimension temporelle du panel est insuffisante. On se propose ici d'utiliser une procédure convergente décrite par CHAMBERLAIN [1984] afin de comparer les résultats avec ceux obtenus en coupe instantanée. Ceci conduit à remettre en question les évaluations généralement admises.

(Stéfan Olivier, 1999, Coût de l'enfant et hétérogénéité individuelle : l'apport des données de panel, *Annales d'économie et de statistique*, n° 54, 1999, 269-291)

mot : lexique scientifique transdisciplinaire.
mot : lexique méthodologique disciplinaire.
mot : langue générale.
mot : terminologie.

Dans ce résumé, plusieurs ensembles lexicaux se dégagent :

1. *Le lexique propre aux écrits scientifiques*, qui renvoie aux procédures ou aux concepts génériques de l'activité scientifique, probablement présents dans nombre d'écrits du même type et dans des disciplines variées : *évaluation, théorique, réaliser des comparaisons, observation (directe), estimation, utiliser une procédure, (obtenir) résultats*. Certains éléments de notre liste sont des collocations (*réaliser des comparaisons, utiliser une procédure*) et doivent être considérés comme des unités lexicales à part entière. Comme on le verra plus loin, le lexique des écrits scientifiques se caractérise par une phraséologie très riche.
2. *Le lexique abstrait non spécialisé*. Ce lexique n'est pas exclusif des écrits scientifiques et apparaît également dans d'autres types d'écrits argumentatifs ou informatifs : *poser un problème, hétérogénéité, la difficulté joue, influence, élément, dimension, conduire à...*
3. *Le lexique méthodologique disciplinaire*. Certains éléments lexicaux « métascientifiques » peuvent être considérés comme disciplinaires ou relevant d'une famille de disciplines : c'est ainsi le cas dans notre exemple d'expressions comme *en coupe instantanée, (comparaison) longitudinale, panel, (réponse) polytomique, procédure (convergente)* qui renvoient aux procédures scientifiques d'un ensemble de sciences sociales comme l'économie ou la démographie – et pour certaines expressions, d'autres sciences appliquées comme la médecine.
4. *Le lexique terminologique*, ici peu spécialisé, renvoie aux objets examinés par l'étude : *coût, ménage, bien-être, aisance financière*.
5. *Le lexique de la langue « générale » ou « commune »* intègre les mots qui ont essentiellement une fonction grammaticale (*l', de, ou, entre, être, ...*) ou, peu spécialisés, ont une probabilité d'occurrence qui ne paraît liée ni à la discipline ni au genre d'écrits (*enfant, préférence, arrivée, départ...*).

Cette première typologie montre qu'il apparaît difficile de dissocier le lexique purement scientifique du lexique abstrait 'à tout faire' qui est particulièrement fréquent dans les écrits scientifiques. La plupart des mots relevés dans notre exemple comme propres aux écrits

scientifiques (*théorique, réaliser des comparaisons*) ne sont par ailleurs pas complètement spécifiques de ce discours et peuvent se rencontrer avec des acceptions légèrement différentes dans d'autres genres textuels. Autrement dit, le lexique transdisciplinaire des écrits n'est pas un lexique spécialisé, mais plutôt un lexique abstrait particulièrement représenté dans les écrits scientifiques. La plupart des chercheurs qui s'essaient à établir des listes finies, généralement pour des besoins didactiques, intègrent ainsi un ensemble de mots abstraits dont un large sous-ensemble n'a rien d'exclusivement scientifique (cf. Phal 1971 ; Pecman 2004 ; Coxhead 2000). Ces chercheurs recourent souvent, pour éviter une trop grande subjectivité, à des méthodes lexicométriques sophistiquées. Phal (1971) et Coxhead (2000) sélectionnent ainsi les mots de forte fréquence dont le caractère transversal est avéré (en ôtant pour Coxhead les mots de très forte fréquence). Drouin (ce numéro) exploite à la fois la méthode des spécificités positives (les mots doivent être spécifiques aux écrits scientifiques par rapport à un corpus de référence journalistique) et de la répartition (les mots doivent apparaître dans 50% des tranches de textes examinés). Les méthodes statistiques ne sont cependant pas infaillibles et les chercheurs sont amenés à remanier les listes établies en ajoutant et en soustrayant au cas par cas quelques éléments.

Les listes de mots simples du lexique scientifique général présentent bien évidemment des limites. La première est liée à la polysémie des lexèmes qui amène certains linguistes comme Peter Blumenthal (ce numéro) à remettre en question la notion même de lexique transversal entre les sciences humaines et les sciences « dures ». Pour lui, le lexique commun aux deux familles scientifiques renvoie soit à des acceptions différenciées (le mot *théorie* en sciences humaines n'aurait pas le sens contraint qu'il rencontre dans les sciences dures), soit à des mots comme *problème* ou *question* qui ne sont pas spécifiques aux écrits scientifiques. Si l'on souhaite s'appuyer sur des référentiels lexicaux pour les applications didactiques, il conviendra donc de prendre en compte autant que faire se peut³ la famille de disciplines, comme le suggèrent Hyland et Tse (2006) qui remettent en cause l'utilité de la liste du lexique académique établie par Coxhead (2000). La deuxième limite des listes établies par les méthodes lexicométriques est la décontextualisation des mots, et la non prise en compte de la phraséologie, qu'il s'agisse d'expressions figées au sens strict, de collocations, voire de formules routinières comme celles que Siepmann (2005 ; ce numéro) intègre dans ses marqueurs de discours. Or, ces expressions polylexicales jouent un rôle de premier plan dans les écrits scientifiques comme on le verra dans la section suivante.

Malgré les limites mentionnées, les inventaires lexicaux de la langue scientifique peuvent apparaître utiles pour certains usages pédagogiques, à condition d'être utilisés à bon escient et recontextualisés. À l'aide de techniques statistiques portant sur les écrits scientifiques, ils permettent de donner un aperçu global des approches épistémologiques véhiculées dans les textes (cf. par exemple la comparaison entre les écrits de lettres et de linguistique effectuée par Rinck (2006) à partir d'une étude du lexique épistémique) ou de proposer des classements textuels (cf. Poudat & Loiseau, ce numéro). Ces listes de vocabulaire sont cependant d'autant plus exploitables qu'elles intègrent des informations sur leur fonction sémantique ou rhétorique, comme on le verra plus loin.

Lexique scientifique, phraséologie et contexte

Le lexique scientifique des écrits scientifiques ne comporte pas que des mots isolés, mais aussi une phraséologie riche qui a fait l'objet de plusieurs travaux de recherche,

³ Averil Coxhead et David Hirsh (ce numéro) soulignent cependant qu'il est souvent difficile, pour des raisons pratiques, de pratiquer un enseignement de l'*English for Academic Purposes* par discipline, les groupes d'étudiants étant hétérogènes.

particulièrement en anglais (cf. Gledhill 2000 ; Howarth 1996 ; Williams 1999 ; Pecman 2004⁴).

Dans une optique didactique ou pour les applications de traitement automatique du langage, il apparaît souvent plus pertinent de se centrer sur les séquences polylexicales que sur les mots simples. Tout d'abord, comme mentionné précédemment, elles présentent moins d'ambiguïté sémantique. Par exemple, Boch, Grossmann & Rinck montrent dans ce numéro qu'un mot comme *conformément* ne renvoie pas en lui-même aux attentes de l'auteur-chercheur et n'aiguille vers ce sens que lorsqu'il est conjoint à des termes comme *hypothèse*, *résultats*, *prévisions*. En outre, les associations lexicales, bien que fréquentes et partiellement motivées sur le plan sémantique (cf. Grossmann & Tutin 2006 ; Tutin 2007b), sont souvent assez difficiles à prédire, ce qui justifie qu'elles soient recensées et traitées de façon systématique, en particulier dans une perspective didactique. Enfin, les séquences polylexicales, qu'il s'agisse de collocations ou d'expressions routinières, remplissent souvent des fonctions rhétoriques centrales dans les écrits scientifiques et sont assez stéréotypées. Cela permet d'envisager une formalisation de ces expressions pour des applications de TAL comme celles qui sont développées par Teufel (1998) ou Sandor (ce numéro)⁵.

Les séquences polylexicales étudiées dans les écrits scientifiques sont assez diverses. Il peut s'agir de collocations restreintes (Tutin 2007a, 2007b ; Drouin, ce numéro) dans la tradition de Hausmann (1989) et de Mel'čuk (1998), c'est-à-dire des associations binaires et récurrentes d'éléments linguistiques entretenant une relation syntaxique comme *résultats encourageants* ou *mener une analyse*. Dans la tradition contextualiste anglaise de Sinclair (1990), Gledhill (2000) donne une définition un peu plus large des collocations (Ex : *it is possible that...*), la cooccurrence plus que la relation syntaxique servant de critère déterminant. La notion est encore plus large chez Williams (1999) pour qui les collocations peuvent être des éléments lexicaux non nécessairement adjacents, ayant simplement tendance à apparaître en cooccurrence. Signalons aussi le travail de Dubreuil (2006) qui montre que, dans les écrits scientifiques, les associations collocationnelles s'avèrent bien souvent argumentativement motivées, et parfois prédictibles. Pecman (2004 ; ce numéro) évite le terme ambigu de collocation et lui préfère le terme générique d'« unité phraséologique » pour les expressions figées, les collocations restreintes et les cooccurrences lexicales plus larges.

Les travaux portant sur les aspects rhétoriques ou pragmatiques dans les écrits scientifiques s'intéressent aussi aux séquences polylexicales dont ils proposent souvent un classement fonctionnel, mais souvent peu structuré sur le plan formel. Teufel (1998) inventorie ainsi des séquences lexicales exprimant des fonctions rhétoriques comme « introduire sa propre solution » (qui peut se traduire linguistiquement par la séquence *the basic idea for the analysis can be seen...*). De la même façon, dans ce numéro, Boch, Grossmann & Rinck proposent une micro-grammaire de la convergence/divergence (*contrairement à nos attentes, comme on pouvait s'y attendre, de façon surprenante...*) en utilisant ces traces lexicales comme indices du modèle de scientificité. Siepmann (2005 ; ce numéro) s'intéresse quant à lui aux marqueurs polylexicaux du discours scientifique qui lient des unités textuelles, et associe à ces relations des fonctions rhétoriques spécifiques, en s'inspirant du modèle de la Rhetorical Structure Theory (Mann & Thompson 1988) ; par exemple, des marqueurs de discours comme *signalons (enfin), pour commencer, reste..., enfin et surtout* ont une fonction d'énumération. Le statut syntaxique accordé aux séquences relevées et analysées n'est cependant pas très formalisé, comme le déplore Sándor dans ce numéro qui propose un

⁴ Soulignons que l'approche de Pecman est contrastive et qu'elle s'intéresse également au français.

⁵ Le projet « Scientext : un corpus et des outils pour étudier le positionnement et le raisonnement de l'auteur sur les écrits scientifiques », financé par l'ANR et piloté par notre laboratoire, se base sur les mêmes présupposés.

modèle plus structuré de « concept-matching » qui allie analyse syntaxique et analyse sémantico-rhétorique.

Les inventaires proposés pour les séquences polylexicales restent dans l'ensemble très fragmentaires, en tout cas pour le français (hormis le travail de Pecman 2004). Pour les applications de TAL et les applications didactiques, il paraît essentiel de proposer une modélisation fine des différents types de phénomènes phraséologiques

Modélisation du lexique

Les listes du lexique scientifique général constituées par Phal ou Coxhead sont brutes. Elles comportent des informations sur les fréquences, et pour le travail de Coxhead, un regroupement par famille morphologique (par exemple, les éléments de la famille de *evaluation* : *evaluate*, *evaluated*, *re-evaluate*, *re-evaluation*...). Pour des exploitations linguistiques fines, il est cependant indispensable de proposer un classement sémantique ou fonctionnel de ce lexique. Cela permettrait, par exemple, un classement des écrits scientifiques selon leur orientation épistémologique, ou de construire des outils de recherche d'information ciblant une fonction rhétorique spécifique, comme le propose Agnes Sándor dans ce numéro. Un classement notionnel (cf. Pecman 2004 et ce numéro) permet aussi d'envisager des applications didactiques comme l'élaboration d'outils d'aide à la rédaction.

Plusieurs types de classements notionnels ont été proposés. Pecman (2004 ; ce numéro) propose une modélisation sémantique fine (la plus aboutie, à notre connaissance) des Unités Phraséologiques de la langue scientifique générale. La description sémantique s'organise autour de quatre sphères notionnelles : la scientificité, l'universalité, la modalité et la discursivité, auxquelles se rattachent des notions comme [EXPERIMENTATION], [RESULTAT] (pour la scientificité). Des étiquettes notionnelles sont rattachées à des ensembles d'expression. Par exemple, l'étiquette [COLLABORATION] renverra aux expressions *développer une collaboration*, *réunir qqn* ou *en collaboration avec*, etc. Ce mode d'organisation devrait permettre un accès onomasiologique au lexique pour des applications d'aide à la rédaction. Pour notre part, nous avons commencé à travailler à l'établissement de classes sémantiques en utilisant des critères linguistiques distributionnels, technique qui présente l'avantage d'être fondée sur des critères reproductibles et partiellement automatisables (Tutin 2007a, 2007b). Les classes obtenues, proposées pour les noms et les verbes, sont également moins abstraites, et peut-être, de ce fait, plus facilement utilisables dans des applications d'aide à la rédaction. Les chercheurs s'intéressant aux aspects rhétoriques et pragmatiques ont également proposé des esquisses de classement du lexique. Teufel (1998) et Sándor (ce numéro) proposent ainsi d'associer à un sous-ensemble de séquences polylexicales des fonctions rhétoriques très spécifiques des écrits scientifiques – par exemple, la fonction « contribution de l'auteur » ou « renvoi aux références de base » – inspirées en partie du modèle CARS (« create a research space » développé par Swales). Moins spécifiques sont les traitements de Hyland (2005) et Dahl (2003) sur les éléments du métadiscours scientifique, distribuant les séquences dans de larges catégories comme le métatexte, les évaluatifs, les marques d'atténuation, les connecteurs, etc. Enfin, d'autres études comme celles de Siepmann (2005 ; ce numéro) contribuent à élaborer des typologies des marqueurs de discours, c'est-à-dire une typologie des fonctions pragmatiques des connecteurs polylexicaux employés dans les écrits scientifiques.

Loin d'être concurrentes, ces différentes approches – sémantiques, pragmatiques, rhétoriques – apparaissent complémentaires et contribuent à une meilleure compréhension des écrits scientifiques. Il nous semble cependant possible, comme le suggère Agnes Sándor dans ce numéro, de mieux articuler les niveaux de traitement sémantique et rhétorique, en associant aux relations sémantiques liant les éléments du lexique des fonctions rhétoriques spécifiques.

Présentation des contributions du numéro

Les deux premières contributions traitent toutes deux du lexique scientifique ou épistémique dans des textes de vulgarisation, les encyclopédies, et les examinent à l'aide de techniques lexicométriques.

Peter Blumenthal compare le lexique nominal commun des articles de sciences humaines et des articles de sciences exactes de l'*Encyclopaedia Universalis* et observe également la combinatoire d'un sous-ensemble de noms. L'examen du lexique partagé et de sa combinatoire révèle de profondes divergences entre les « deux cultures ».

Céline Poudat et Sylvain Loiseau s'intéressent à l'encyclopédie collaborative *Wikipédia*. Ils décrivent le contenu de l'encyclopédie sur le plan linguistique, en se focalisant sur les sciences et leur représentation. L'étude, fondée sur un ensemble de corpus extraits à partir de l'arborescence des catégories de l'encyclopédie, montre des résultats contrastés sur les plans lexicaux général et épistémique.

Une seconde série d'articles s'attaque à la question de la définition concrète du lexique scientifique général, dans une perspective didactique, en utilisant les outils de la linguistique de corpus.

Patrick Drouin cherche à isoler le lexique scientifique transdisciplinaire, ce lexique qui serait commun à toutes les disciplines, en utilisant des techniques statistiques, dont le calcul des spécificités, sur un gros corpus de thèses. L'approche employée apparaît pertinente, même si un filtrage manuel s'avère indispensable.

Averil Coxhead et David Hirsh partent du constat que la liste du lexique académique élaborée par Coxhead pour l'*English for Academic Purposes (Academic Word List)* ne permet pas de traiter une partie des écrits de sciences exactes et appliquées. À l'aide de corpus diversifiés, ils élaborent une liste de mots purement « scientifiques » dont la couverture apparaît satisfaisante.

Mojca Pecman propose une méthode permettant d'explorer les invariants des différents types de discours scientifiques, qui combine plusieurs approches : transdisciplinaire, onomasiologique et phraséologique. Une analyse notionnelle et combinatoire est proposée pour le phraséolexique de la langue scientifique générale.

La dernière série de contributions aborde le lexique à travers les questions rhétoriques et pragmatiques.

Agnes Sándor propose un modèle, le « concept-matching » permettant de formaliser les éléments du metadiscours et de leur associer une fonction rhétorique. Ce modèle, implémenté dans un système de traitement automatique de langues naturelles, permet de détecter le metadiscours particulier qui véhicule la stratégie rhétorique de l'auteur.

Françoise Boch, Francis Grossmann et Fanny Rinck proposent une analyse qualitative des marqueurs lexicaux de la convergence (de type *conformément à nos hypothèses*) et de la divergence (de type *contre toute attente*) par rapport aux attentes du chercheur-auteur ou de la communauté scientifique. L'étude montre que si ces marqueurs ne présagent pas de la démarche scientifique impliquée dans l'ensemble de l'article, ils peuvent en revanche révéler des styles épistémologiques.

Dirk Siepmann propose une définition opérationnelle et une typologie des marqueurs du discours polylexicaux selon des critères fonctionnels. Il met l'accent sur les marqueurs de suggestion, et met en évidence que l'abondance de ces marqueurs discrédite l'hypothèse selon laquelle la langue des sciences serait libre de toute subjectivité.

Remerciements. Un grand merci à Françoise Boch qui a relu une version préalable de cette présentation.

Agnès Tutin
 Université Grenoble 3
 Laboratoire de Linguistique et Didactique des Langues Etrangères et Maternelles (LIDILEM),
 BP 25, 38040 Grenoble Cedex 09
 <Agnes.Tutin@u-grenoble3.fr>

Références

- Biber, D. (2006). *University Language: a corpus-based study of spoken and written registers*. Amsterdam, Benjamins.
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34 (2), 213-238.
- Dahl, T. (2003). Metadiscourse in research article. In Fløttum, K., Rastier F. (eds). *Academic Discourse. Multidisciplinary Approaches*, Oslo, Novus Press.
- Dubreuil, E. (2006). *La dimension argumentative des collocations textuelles en corpus électronique spécialisé dans le domaine du TAL(N)*. Thèse de doctorat, Université de Nantes.
- Fløttum, K. (ed.) (2007). *Language and Discipline Perspectives on Academic Discourse*. Cambridge, Cambridge Scholars Publishing, Newcastle.
- Fløttum, K., Dahl, T. & Kinn, T. (2006). *Academic Voices. Across languages and disciplines*. Amsterdam, Benjamins.
- Gledhill, Ch. (2000). *Collocations in Science Writing*. Language in performance, 22. Tübingen, Gunter Narr Verlag.
- Grossmann, F. & Tutin, A. (2007). Motivation of lexical associations in collocations: the case of intensifiers for “nouns of joy”. In Wanner L. (ed.), *Selected Lexical and Grammatical Issues in the Meaning-Text Theory In honor of Igor Mel'čuk*, Amsterdam, Benjamins, 139-165.
- Grossmann, F. & Wirth, F. (2007). Marking Evidentiality in Scientific Papers: The Case of Expectation Markers. In K. Fløttum (ed.), *Language and Discipline Perspectives on Academic Discourse*. Cambridge, Cambridge Scholars Publishing, Newcastle, 202-218.
- Harris, Z., Ryckman, T., Harris, S., Gottfried, M., Daladier, A., Mattick, P. (1989). *The Form of Information in Science, Analysis of Immunology Sublanguage*. Dordrecht, Kluwer.
- Hausmann, F.J. (1989). Le dictionnaire de collocations. In Hausmann F.J., Reichmann O., Wiegand H.E., Zgusta L. (eds.), *Wörterbücher: ein internationales Handbuch zur Lexicographie. Dictionaries. Dictionnaires*. Berlin/New-York, De Gruyter. 1010-1019.
- Howarth, P.A. (1996). *Phraseology in English Academic Writing*. Tübingen, Niemeyer.
- Hyland, K. & Tse, P. (2007). Is there an ‘academic vocabulary’? *TESOL Quarterly*, 41, 235-253.
- Hyland, K. (1998). *Hedging in Scientific Research Articles*. Amsterdam, Benjamins.
- Hyland, K. (2005). *Metadiscourse*. London-New York, Continuum.
- Hyland, K. & Bondi, M. (eds.) (2006). *Academic Discourse Across Disciplines*. Bern, Peter Lang.
- Mann, W.C. & Thompson, S.A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8 (3), 243-281.
- Mel'čuk, I. (1998). Collocations and Lexical Functions. In A.P. Cowie (ed.), *Phraseology. Theory, Analysis and Applications*, Oxford, Clarendon Press.
- Pecman, M. (2004). *Phraséologie contrastive anglais-français : analyse et traitement en vue de l'aide à la rédaction scientifique*. Thèse de doctorat, Université de Nice-Sophia Antipolis.
- Phal, A. (1971). *Vocabulaire général d'orientation scientifique (VGOS) - Part du lexique commun dans l'expression scientifique*. Paris, Didier.
- Poudat, C. (2006). *Etude contrastive de l'article scientifique de revue linguistique*. Thèse de doctorat, Université d'Orléans.
- Rinck, F. (2006). *L'article de recherche en Sciences du Langage et en Lettres, Figure de l'auteur et approche disciplinaire du genre*. Thèse de doctorat, Université de Grenoble.

- Rinck, F., Boch, F., Grossmann, F. (2007). Quelques lieux de variation du positionnement énonciatif dans l'article de recherche. In Lambert, P., Millet, A., Rispail, M. & Trimaille, C. (eds.), *Variations au coeur et aux marges de la sociolinguistique*, Paris, L'Harmattan, Espaces Discursifs, Paris.
- Siepmann, D. (2005). *Discourse Markers across Languages*. Londres, Routledge.
- Swales, J.M. (1990). *Genre Analysis: English in academic and research settings*. Cambridge, Cambridge University Press.
- Teufel, S. (1998). Meta-discourse markers and problem-structuring in scientific articles. In M. Stede, L. Wanner, E. Hovy (eds.), *ACL 1998 Workshop, Discourse Structure and Discourse Markers*. Montréal, Somerset/New Jersey, 43-49.
- Tutin, A. (2007a). Traitement sémantique par analyse distributionnelle des noms transdisciplinaires des écrits scientifiques. *Actes de TALN 2007. Communications affichées*, 283-292.
- Tutin, A. (2007b). Modélisation linguistique et annotation des collocations : application au lexique transdisciplinaire des écrits scientifiques. In S. Koeva, D. Maurel, M. Silberztein (eds.), *Formaliser les langues avec l'ordinateur*, Besançon, Presses Universitaires de Franche-Comté.
- Williams, G.C. (1999). *Les réseaux collocationnels dans la construction et l'exploitation d'un corpus dans le cadre d'une communauté de discours scientifique*. Thèse de doctorat, Université de Nantes.